Generative Flows on Discrete State-Spaces: Enabling Multimodal Flows with Applications to Protein Co-Design

Year: 2024 | Citations: 208 | Authors: Andrew Campbell, Jason Yim, R. Barzilay, Tom Rainforth, T. Jaakkola

Abstract

Combining discrete and continuous data is an important capability for generative models. We present Discrete Flow Models (DFMs), a new flow-based model of discrete data that provides the missing link in enabling flow-based generative models to be applied to multimodal continuous and discrete data problems. Our key insight is that the discrete equivalent of continuous space flow matching can be realized using Continuous Time Markov Chains. DFMs benefit from a simple derivation that includes discrete diffusion models as a specific instance while allowing improved performance over existing diffusion-based approaches. We utilize our DFMs method to build a multimodal flow-based modeling framework. We apply this capability to the task of protein co-design, wherein we learn a model for jointly generating protein structure and sequence. Our approach achieves state-of-the-art co-design performance while allowing the same multimodal model to be used for flexible generation of the sequence or structure.